DISCUSSION PAPER SERIES

IZA

IZA DP No. 3204

Migrant Networks, Migrant Selection, and High School Graduation in Mexico

Alfonso Miranda

December 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Migrant Networks, Migrant Selection, and High School Graduation in Mexico

Alfonso Miranda
Keele University
and IZA

Discussion Paper No. 3204
December 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 3204
December 2007

ABSTRACT

Migrant Networks, Migrant Selection, and High School Graduation in Mexico*

This paper examines whether family and community migration experience affect the probability of high school graduation in Mexico once unobserved heterogeneity is accounted for. Bivariate random effects dynamic probit models for cluster data are estimated to control for the endogeneity of education and migrant network variables. Correlation of unobservables across migration and education decisions as well as within groups of individuals such as the family are explicitly controlled for. Results show that migrant networks reduce the likelihood of high school graduation. Negative migrant selection is detected at the individual level while positive migrant selection is found at the family level.

JEL Classification: F22, I21, J61, C35

Keywords: migration, education, migrant selection, dynamic bivariate probit

Corresponding author:

Alfonso Miranda
School of Economic and Management Studies
Keele University
Keele, Staffordshire ST5 5BG
United Kingdom
E-mail: A.Miranda@econ.keele.ac.uk

* I am grateful to Christopher Woodruff for kindly providing data on historical migration and state distance to the nearest station on the north/south rail lines in Mexico in the early 1900s. I am also grateful to Gauthier Lanot, Wiji Arulampalam, Mark Stewart, Massimiliano Bratti, Leslie Rosenthal, Amanda Gosling, João Santos Silva, attendants to the 2007 annual migration meeting, and seminar attendants at Keele and Kent for useful comments.

1 Introduction

In the last few decades international migration has become a topic of primary interest. The main destinations, Northern America and Europe, received nearly 13.1 million of new immigrants between 2000 and 2005. In contrast, Asia, Latin America and the Caribbean, the main origin areas, sent 10.1 million emigrants during the same five-year period (UNPD 2006). This intensive international flow of labour creates a number of economic, political, and social challenges that are attracting more and more the attention from policy makers and international organisations.

Traditionally, academic research in the field has focused in understanding the effects that immigration has on the labour market of the host country. In recent years, however, there is an increasing interest in learning whether international migration has impacts on poverty, accumulation of human and physical capital, economic growth, and development in source countries (see, for instance, World Bank 2006). The Mexico-US is a leading case of interest because in the last two decades the flow of labour from Mexico to the US reached unprecedented numbers.[1] In addition, the amount of remittances sent by Mexican expatriates to their families in Mexico increased steadily and became, without doubt, a non-negligible source of income. In fact, Banco de México estimates that in 2005 remittances from the US represented nearly 2.6% of the GDP of Mexico.

The present paper intends to contribute a study on these issues. In partic-

[1] Mexico is by far the main origin country in Latin America. In fact, during the period 2000–2005 alone, Mexico sent nearly 2 million of emigrants to the United States (UNPD 2006).

ular, attention is focused on learning whether Mexico-US migration networks affect the likelihood of high school graduation in Mexico.

Econometric work is essentially complicated by the fact that individual unobserved heterogeneity affecting migration choice is potentially correlated with unobserved traits affecting education decisions. Unobserved skills are a good example. On the one hand, the labour economics literature stresses the fact that skilled individuals are more likely to succeed at school and to find qualified jobs (see, for instance, Miranda and Bratti 2006, Blundell et al. 2000). On the other hand, the migration literature points out that returns to education are higher in Mexico than in the US and that unqualified jobs are better paid in the American side. As a consequence, Mexican unskilled workers have strong incentives to emigrate to the US (see, for instance, Borjas 1994). A negative correlation among unobservables is therefore expected because skilled individuals are likely to study more and emigrate less. Clearly, failing to account explicitly for such a correlation may be a cause of serious bias.

Besides correlation across education and migration choices at the individual level, unobservables can be correlated within certain groups of individuals. The family is an obvious unit for this type of clustering because members of a kin share a set of unobservable traits (say, for instance, genetic make-up or common adverse shocks) that affect their performance at school and change their likelihood of migration. Failing to account for this "intra-family clustering" can lead, once again, to serious bias.

Controlling for intra-family clustering is also important because an individual's education and migration decisions can be a function of the choices taken by other members of her/his kin. For instance, individuals who are successful at school can create peer pressure and learning resources for other members of their family and influence their school performance. Similarly, a migrant individual can help his/her relatives to migrate (i.e., to access migrant network resources) and/or to provide them with a successful role model of migration (i.e., access information and 'reputation' spillovers). Finally, important dynamic cross-effects may be present because, if an individual migrates, other members of her/his kin can benefit from the money she/he sends home and from the contacts he/she builds up at the destination country in such a way that their incentives to study in the home country are changed. However, unless sources of unobserved variation at the individual and family level are set apart, the researcher will find impossible to distinguish between real and spurious dynamic dependence — the latter being dynamic dependence induced by unaccounted unobserved heterogeneity (see Arulampalam and Bhalotra 2006, Heckman 1981a). Further, if unobserved heterogeneity is not properly controlled for, inconsistent estimators can be obtained because family migrant/education network variables are likely to be endogenous.

The present paper addresses all these econometric challenges by estimating bivariate random effects dynamic probit models for cluster data.

To date the literature has not fully recognised the complexity of the relationship between migration and education. There are two main strands of study. One strand is related to the analysis of social networks and its influence on migration decisions (see, for instance, Delechat 2001, Winters et al. 2001). These studies commonly use univariate dynamic probit models to disentangle the effects of migrant networks and previous migration experience on current migration events. Unfortunately, correlation of unobservables across migration and educational outcomes is not allowed. The second strand is concerned directly with the effects of migration networks on educational attaintment in origin countries and has produced relatively fewer pieces of work. In this category are McKenzie and Rapoport (2006), and Hanson and Woodruff (2003). These authors use instrumental variable techniques to control for the correlation of unobservables across migration and schooling variables. However, none of them allow for intra-family clustering. To the knowledge of the author, no previous study has addressed both potential problems simultaneously.

The study uses data from the Mexican Migration Project (MMP). The MMP is a rich individual-level data set that contains detailed information on migrant networks and collects information about the head of the household and all her/his sons and daughters independently of the current location of the latter individuals — therefore, long-term emigrants are well covered. Further, legal and illegal border crossings are carefully recorded.

Results suggest that family and community migration networks decrease significantly the likelihood of high school graduation in Mexico. In particular, it is found that an extra migrant in the family decreases the likelihood of high school graduation by 2.4 percentage points. Similarly, results show that if a community is an extra 1000 kilometres from areas with long-lasting migrant tradition, the likelihood of high school graduation increases as much as 17.4 percentage points. Regarding migration, empirical evidence suggests that having a migrant in the family increases by 4 percentage points the chances of observing the next younger member migrating as well. The self-enhancing effects of family migrant networks are, however, limited. In fact, evidence shows that as the number of migrants in the family accumulate, the remaining members become less likely to migrate themselves. The study also finds evidence that community migrant networks increase the odds of observing a new migration event. In particular, evidence shows that being an extra 1000 kilometres closer from areas with long-lasting migrant tradition increases the likelihood of migration by 9 percentage points.

Evidence of negative migrant selection in terms of unobservables is detected at the individual level while positive migrant selection is detected at the family level. The finding may help to bring together apparently contradicting evidence on migration selectivity put forward by Borjas (1994) and Chiquiar and Hanson (2005).

2 Do migrant networks affect education? Why?

When migrants leave their home country family and friends are commonly left behind. Once established at the destination, migrants keep close contact with their communities back home and, in many cases, send money (remittances) and help members of their kin to migrate themselves.

The money migrants send home is used in a number of ways, including helping credit-constrained individuals to achieve their desired level of education. This option is particularly attractive to those who have no plans to emigrate themselves and education offers them an opportunity to improve their standard of life at the home country. As a consequence, through its links with remittances, migrant networks are expected to increase education at the source county (see, for instance, World Bank 2006).[2]

The story, however, does not end there. A group of recipient individuals plan to leave the home country. Those individuals will use remittances to finance their education at the source if observable qualifications are broadly portable across host and source countries (for more on this argument, see Vidal 1998).[3] In contrast, if observable qualifications are non-portable, rational prospective migrants will behave in a forward looking fashion and drop out school early to avoid wastage of valuable resources (a similar argument is put forward by McKenzie and Rapoport 2006). Finally, if qualifications are 'noisily' portable, then a zero effect of migrant networks on education at the source country may be observed.

Even if migrants do not send money home they can still affect education decisions at the source country. Namely, through their networks, current migrants can help members of their kin and/or community to migrate and to reduce labour market uncertainties at the destination country. Access to the resources of a migrant network can, once again, incentive prospective migrants to behave in forward looking fashion and drop out school early if observable qualifications are non portable across borders. On the contrary, if qualifications are portable, the opposite effect may be observed.

Clearly, from a purely theoretical point of view, the effect of migrant networks on education in the source country is ambiguous even if one is

[2]Obviously, a zero effect can be observed either because households are not credit-constrained in the first place, or because the contribution of remittances do not change significantly the overall financial position of recipient households.

[3]Under such an assumption acquiring education at the origin country is an efficient way to improve the odds of a highly paid job at the destination country. This route will be attractive specially if prospective migrants have no access to education at the destination country.

willing to assume that qualifications are non-portable across borders. For these reasons, empirical investigation is needed.

3 Data

Data from the Mexican Migration Project (MMP) are used. The MMP is a pooled cross section of migrant communities located throughout Mexico, which is collected by a joint group of researchers at Princeton University and Universidad de Guadalajara.[4] Every year, from 1982 to 2005, members of the MMP team survey a random sample of 200 households in two to five communities in Mexico to gather a new cross-section. Such cross-section is then added to the pool. Current files, the MMP107 database, contain information at individual and community level in 107 localities.

The communities surveyed by the MMP are not selected at random. As a consequence, the data may not be argued to be National or State representative. Instead, the MMP107 is representative of the population in the 107 communities that are included in the study. Very importantly, selected communities are chosen on the basis that they have some, though not necessarily long-lasting, migrant tradition. Across the years, the MMP team has managed to survey communities in many regions of the country and with different sizes, from small rural towns to large cities.[5] Moreover, there has been some effort to select communities so that there is enough variation in terms of economic activity — from small places that specialise in mining,

[4]Data files are freely available at http://mmp.opr.princeton.edu/

[5]Twenty States are covered: Aguascalientes, Baja California Norte, Chihuahua, Colima, Durango, Guanajuato, Guerrero, Hidalgo, Jalisco, Michoacan, Nayarit, Nuevo Leon, Oaxaca, Puebla, San Luis Potosi, Sinaloa, Tlaxcala, Veracruz, and Zacatecas.

fishing, and farming, to large urban areas that are highly diversified.

National representative surveys commonly contain too few observations of migrant individuals to allow meaningful statistical analysis (CONAPO 2000). As a consequence, there is always a need to over-sample areas with migrant tradition if useful numbers of migrants are to be obtained. Moreover, it is well-documented that migrants do not come at random from all the geographical areas of Mexico. Instead, they cluster intensively in the States and areas covered by the MMP107 (CONAPO 2000). Hence, if a trend is not present in the MMP107 data, it will hardly appear in a national representative survey. From this point of view, using the MMP107 to perform exploratory analyses of Mexico-US migration issues is well justified and a number of influential papers in the field have used the survey (see, for instance, Delechat 2001, Durand et al. 1996).

The MMP107 has characteristics that made it an important source of information for the study of migration. First, and substantively, it is the only Mexico-US migration survey that covers long-term migrants. In particular, information about the head of the household and all her/his sons and daughters is gathered, independently of the current location or household membership status of the latter individuals. This implies that data for all sons and daughters is available even if some of them formed their own households and emigrated to the US — and haven't come back — many years before the survey. Further, an individual's emigration event is recorded regardless of her/his legal status in the United States. Date and destination of every legal or illegal border crossing in an individual's life history is carefully documented. The other two major surveys about Mexico-US migration, the

ENADID and the MXLFS, do not cover long-term emigrants.[6]

The present study is based on information for 11,990 individuals collected in 3,479 families in 36 rural and urban communities throughout Mexico between 1997 and 2004.[7] Individuals are clustered in families. Within a family the MMP107 gives information about the age of each member.

Since the focus of the paper is high school graduation, only individuals aged 18 or over at the time of the survey are included in the sample. The MMP107 contains information on whether individuals have ever migrated to the US (usmigra=1) and on whether they graduated from high school (prepa=1). These are the two dichotomous dependent variables. Eighteen per cent of the individuals have migrated to the US at least once. Similarly, twenty one per cent of the sample are high school graduates. Migrants are clearly less educated. In fact, 12% of the migrants are graduates compared to the 22% of non-migrants. Table 1 contains summary statistics.

[Table 1 around here]

4 Econometric Issues

Dynamic bivariate random effects Probit models for cluster data are used for the analysis.[8] Denote by M_{ji} the variable that takes on one if, by the

[6] In both cases information is collected for persons who lived in the household up to five years before the date of the survey. Anyone who left the household before that is not considered a member and no information is recorded. This is unfortunate because, most likely, many migrants do not comply with such requirements.

[7] In previous years the MMP survey did not collected data for some of the relevant variables for the analysis. For these reasons, the present study uses data gathered from 1997 onwards.

[8] The model outlined in the present section should be seen more in the tradition of multi-level modelling, a broadly used methodology in statistics, than in the economics tradition

time of the survey, the i-th member of the j-th family has emigrated to the US at least once and zero otherwise. Similarly, E_{ji} indicates whether the i-th member of the j-th family graduated from high school ($E_{ji} = 1$) or not ($E_{ji} = 0$) by the time of the survey. Within families, and to avoid any ambiguity, heads are given the $i = 0$ index and their spouse are given the $i = 1$ index. Children of the head are then ordered by age. The data is, therefore, ordered in such a way that the jk-th individual is older than the jl-th whenever $l > k$. Families can have a different number of members — including the singleton family — and the head may or may not be in a union. Therefore, some families are composed by head, spouse, and children while others have head and children or only head.[9]

4.1 Dynamic equations

A latent variable framework is the natural approach. Let M^*_{ji} and E^*_{ji} be two latent continuous variables. The econometrician does not observe M^*_{ji}

of panel data analysis. Multilevel methods emphasise the need to correctly model data that have a hierarchical (nested or clustered) structure exploiting informative within-cluster variation. Similarly, analysis of panel data intends to correctly model longitudinal data (i.e., data where an individual is observed at least twice) exploiting informative within-individual variation. These two traditions have many points of contact. In fact, if one acknowledges that in longitudinal data measurement occasions (level 1) are clustered within individuals (level 2), it is clear that most panel methods used in econometrics are multilevel models with a random intercept. From this point of view, multilevel modelling is a more general approach. For an excellent review on multilevel modelling see, for example, Goldstein (2003) and McCulloch and Searle (2001).

[9] Giving the $i = 2$ index to the spouse guarantees that all spouses are treated symmetrically in the family independently of the age of the head's children. This addresses some potential problems that may arise from the fact that some heads may have had previous unions and that some of the head's children may be older than the current spouse. Preliminary regressions showed that ordering children and spouse by age irrespectively of their different family role had no significant impact on the empirical results. Performing the analysis only on intact families (those who have head, spouse, and children) also reported similar results.

and E^*_{ji}. Instead two dichotomised variables, M_{ji} and E_{ji}, are available. It is supposed that the high school dummy is generated according to the following data generating process,

$$E^*_{ji} = \mathbf{x^e_{ji}}\boldsymbol{\beta^e} + \delta_{11}E_{j,i-1} + \delta_{12}M_{j,i-1} + \pi_{11}MS_{ji} + \pi_{12}ES_{ji} + f^e_j + u^e_{ji}, \quad (1)$$

with $E_{ji} = 1$ if $E^*_{ji} > 0$ and zero otherwise, $i = \{1\ldots, I\}$ and $j = \{1\ldots, J\}$. Notice that the initial observation within the j-th family is indexed by zero, $i = 0$. Vector $\mathbf{x^e_{ji}}$ represents a set of observed characteristics that can vary at the individual, family, and community levels. Elements of $\mathbf{x^e_{ji}}$ are assumed to be strictly exogenous and $\boldsymbol{\beta^e}$ denotes a conformable coefficient vector — including the constant term. Similarly, $\delta_1 = \{\delta_{11}, \delta_{12}\} \in \mathbb{R}^2$ represent coefficients on the migration and education outcomes of the immediately elder family member. Variables MS_{ji} and ES_{ji} represent the stock of migrants and high school graduates accumulated in the family up to the i-th individual and $\pi_1 = \{\pi_{11}, \pi_{12}\} \in \mathbb{R}^2$ are its associated coefficients. Hence, MS_{ji} and ES_{ji} are a measure of an individual's access to family migration and education networks net of the effect of the outcomes of the immediately older member of the family.[10] Finally, variables f^e_j and u^e_{ji} are random heterogeneity terms. One term, f^e_j, varies at the family level while the other term, u^e_{ji}, varies at

[10]For all heads $MS_{j0} = 0$ and $ES_{j0} = 0$. If $M_{j0} = 1$ ($E_{j0} = 1$) then $MS_{j1} = 1$ ($ES_{j1} = 1$) and zero otherwise. In this fashion, MS_{ji} (ES_{ji}) is increased by one unit every time the $(i-1)$-th individual in the j-th family is observed to migrate (graduated from high school). Hence, MS_{ji} and ES_{ji} are the running cumulative sums of the dependent variables M_{ji} and E_{ji} for $i > 0$.

the individual level. The equation for the migration dummy is,

$$M^*_{ji} = \mathbf{x^m_{ji}}\boldsymbol{\beta^m} + \delta_{21}E_{j,i-1} + \delta_{22}M_{j,i-1} + f^m_j + \pi_{21}MS_{ji} + \pi_{22}ES_{ji} + u^m_{ji}, \quad (2)$$

with $M_{ji} = 1$ if $M^*_{ji} > 0$ and zero otherwise. Following Alessie et al. (2004), f^m_j and f^e_j are specified to be jointly Normally distributed with mean vector zero and covariance matrix Σ_f,

$$\Sigma_f = \begin{bmatrix} \sigma^2_m & \rho\,\sigma_m\sigma_e \\ \rho\,\sigma_m\sigma_e & \sigma^2_e \end{bmatrix}.$$

In a similar fashion, u^m_{ji} and u^e_{ji} are jointly Normal with mean vector zero and covariance matrix Σ_u,

$$\Sigma_u = \begin{bmatrix} 1 & \rho_u \\ \rho_u & 1 \end{bmatrix}.$$

To close the model it is assumed that f^h_j and u^h_{ji} are independent, for $h = (m, e)$. Further, errors f^h_j and u^h_{jk} are serially uncorrelated for every j and k.

The model implies the following relationships. M^*_{ji} and M^*_{jk}, $k \neq i$, are correlated within the j-th family through the random term f^m_j. However, no such correlation exist among individuals who belong to different families. Intra-family clustering is also induced between E^*_{ji} and E^*_{jk} by the random term f^e_j. Also, at the family level, correlation between E^*_{ji} and M^*_{jk} for all i and k that belong to the j-th family is induced by correlation between f^e_j

and f_j^m. Finally, at the individual level, correlation between M_{ji}^* and E_{ji}^* is created by correlation between u_{ji}^m and u_{ji}^e. True dynamic dependence is present if at least one element of vectors $\delta = (\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22})$ and/or vector $\pi = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$ is different from zero.

4.2 Initial conditions

Given that migration and education outcomes of different members within the j-th family are correlated, treating M_{j0} and E_{j0} — and therefore SM_{j1} and SE_{j1} — as exogenous in system (1)-(2) will produce inconsistent estimators. This is known in the econometrics literature as the initial conditions problem. To address the problem we follow the strategy suggested by Heckman (1981b). Namely, a model for the reduced-form marginal probability of M_{j0} and E_{j0} given f_j^e and f_j^m is specified. Hence two further equations are needed,

$$E_{j0}^* = \mathbf{z_{j0}^e}\boldsymbol{\gamma^e} + \lambda_{11} f_j^e + \lambda_{12} f_j^m + v_{j0}^e \tag{3}$$

$$M_{j0}^* = \mathbf{z_{j0}^m}\boldsymbol{\gamma^m} + \lambda_{21} f_j^e + \lambda_{22} f_j^m + v_{j0}^m \tag{4}$$

with $E_{j0} = 1$ if $E_{j0}^* > 0$ and $M_{j0} = 1$ if $M_{j0}^* > 0$ and zero otherwise. As usual, $\mathbf{z_{j0}^e}$ and $\mathbf{z_{j0}^m}$ represent vectors of explanatory variables that can vary at the individual, family, and community level. Notice that migration and education stock variables MS_{ji} and ES_{ji} are excluded from the initial conditions equations as they are, by definition, zero for all heads of family. Coefficients $\lambda = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}) \in \mathbb{R}^4$ represent free parameters (*factors loadings*) that allow any type of correlation among E_{j0}^*, M_{j0}^*, E_{ji}^*, and M_{ji}^*. We suppose that v_{j0}^h is uncorrelated with v_{jk}^h for every j and k. As usual, v_{j0}^e

and v_{j0}^m are jointly normal with mean vector zero and covariance matrix Σ_v,

$$\Sigma_v = \begin{bmatrix} 1 & \rho_v \\ \rho_v & 1 \end{bmatrix}.$$

4.3 Identification and control variables

Technically the model is identified through functional form (see Heckman 1978). However, in the absence of exclusion restrictions identification may be 'tenuous' (in the context of the multinomial probit model see Keane 1992). Hence, specifying exclusion restrictions to help identification is a good practise.

Using information from the MMP survey one can identify the main US city/urban area destination of each community in the sample between 1990 and 2000. Similarly, local area unemployment rates and labour force statistics in the US are available from the Bureau of Labor Statistics (BLS). Hence, it is possible to obtain an average unemployment rate (laur) and size of the labour force (lforce) between 1990 and 2000 for each local area reported by the BLS and match such information with the MMP data. Both laur and lforce are indicators of the labour market characteristics of the main US city/urban area destination of the MMP communities included in the sample.

Variables laur and lforce enter the migration equations but are excluded from the schooling equations. Clearly, unemployment rate at the community's main US destination is a good indicator of how difficult is for new immigrants to find a job at arrival. The higher laur is the less attractive migration will be for prospective migrants. Similarly, large cities have complex economies and are more capable of absorbing people with different skills and backgrounds than small urban areas. As a consequence, one can expect migration to be more attractive as lforce becomes larger. Both laur and lforce are unlikely to affect education decisions in Mexico and, if they do, it is exclusively through their impact on migration. These two variables are, therefore, good candidates for imposing exclusion restrictions to help identification.

Following Woodruff and Zenteno (2007) community migrant networks are controlled for by exploiting variation in an individual's degree of access to historical migrant networks. Given that historical migration may be endogenous in system (1)-(4), a proxy that is unlikely to be correlated with all $f^{(\cdot)}$, $u^{(\cdot)}$, and $v^{(\cdot)}$ is used instead. In particular, access to community migrant networks is approximated by the distance from the capital of the state in which an individual was born to the nearest station on the north/south rail lines in the early 1900s, raildis.[11]

As indicated by Woodruff and Zenteno, the rationale behind the use of raildis as a proxy for historical migration and, therefore, as a proxy for current access to community migrant networks is that during the first two decades of the 20th century a large number of Mexican workers were recruited to work in the south of the US. Given the lack of important population centres nearby the border and other more efficient means of transport, American contractors went down the existing Mexican north/south railway route hiring mexican

[11] Woodruff and Zenteno calculate raildis as the distance from the capital city of each state to a stop on any of the main north/south rail lines as they existed in the early 1900s. Where the line passed through the state, as is the case in 16 states, a distance of zero was assigned. For border states not served by the rail line and for Baja California Sur, raildis is the distance from the capital city to the border. I thank the authors for providing me with these data.

citizens along the way. Also, between 1910 and 1921, the north/south railway played a central role in moving troops during the Mexican revolution. When either the *villista* or the federal army went trough a population nearby the railway, they would stop to get supplies and force young men in town to join the army (see Taibo II 2006, for an excellent narrative and extensive historical reference of the movement of troops along the rails during the Mexican revolution and the implications for towns, villages, and cities that were nearby). As a response to this menace, many men and families fled to the north using the railway as soon as they have news it was safe to do so. Part of this displaced population went as far as to the south of the US.

Proximity to the railway in early 1900s in Mexico gave, therefore, reasons and opportunities to Mexican citizens to migrate to the US. These early movements of labour and displaced population helped to accumulate experience and contacts in the US that were used upon the end of the second world war to send new waves of migrants. There are, as a consequence, good reasons to suggest that raildis is correlated with current access to community migrant networks. In contrast, no reasons lead us to believe that raildis should be correlated with other current community characteristics that may affect education and migration in the present day. One can reasonably sustain, therefore, that raildis is neither correlated with family level random effects $f^{(\cdot)}$ nor with individual level error terms $u^{(\cdot)}$ and $v^{(\cdot)}$. Hence, it is reasonably justified to use raildis as a proxy for access to community migrant networks in system (1)-(4).

Other explanatory variables include sex, age, total number of children the head ever had. Dummies for rural/urban classification of the surveyed communities as well as birthplace, region, and survey year are also included. Variables such as income, labour participation status, and wage are not included in the list of explanatory variables because it are very likely to be endogenous and the MMP does not contain information on characteristics than can be used as valid instruments. Hence, equations (1)-(4) should be seen as a reduced-form model and the reader should have the due care when interpreting results.

4.4 Estimation strategy

The model is estimated by Maximum Simulated Likelihood (see, for instance, Train 2003). The contribution of the j-th family to the likelihood is,

$$\begin{aligned} L &= \int\int \Phi_2\left(q_1 w_{11}, q_2 w_{12}, q_1 q_2 \rho_v\right) \\ &\quad \times \prod_{j=1}^{J} \Phi_2\left(q_1 w_{21}, q_2 w_{22}, q_1 q_2 \rho_u\right) g\left(f^e, f^m, \Sigma_f\right) df^e df^m \end{aligned} \tag{5}$$

where $g(.)$ represents the bivariate normal density of the family random effects, $q_1 = 2E_{ji}-1$, and $q_2 = 2M_{ji}-1$. Finally, w_{11} and w_{12} are the right-hand side of equations (3) and (4) excluding u_{ji}^e and u_{ji}^m respectively. Variables w_{21} and w_{22} are defined in the same fashion using equations (1) and (2).

Two uncorrelated Halton sequences of dimension R are first obtained. Then random draws from density $g(.)$ are simulated using the Halton sequences, a Cholesky decomposition, and the inverse cumulative normal distribution. Next, for each draw (which is a two dimension vector), the conditional likelihood of the j-th family is evaluated. Finally, an average of the R simulated conditional likelihoods is taken. This average is the contribution of the j-th family to the overall simulated likelihood — an approximation of the double integral in (5). Halton sequences have been shown to achieve high precision with fewer draws than uniform pseudorandom sequences because it have a better coverage of the $[0, 1]$ interval (for more on this topic see Train 2003).

Maximum simulated likelihood is asymptotically equivalent to ML as long as R grows faster than $\sqrt{N}$ (Gourieroux and Monfort 1993). Following Alessie et al. (2004) maximisation is performed on the basis of the BHHH algorithm. At convergence, numerical second derivatives are obtained to calculate the robust covariance matrix.

5 Empirical Results

Table 2 presents the results. For comparison reasons, table 2 contains results from univariate dynamic probit models for usmigra and prepa along with the estimates from the bivariate dynamic model. Regressions were initially estimated using 200 Halton draws. Then, 50 draws were successively added until no significant differences in coefficients and log-likelihood were detected. In all cases 400 Halton draws were enough to achieve high precision. Marginal effects (MEs) are calculated at the means of the independent variables and robust standard errors are obtained using the delta method.

[Table 2 around here]

Let us start the discussion with the results from the univariate models. In either case, prepa and usmigra, there is evidence of non-negligible intra-family clustering as the estimates for σ_e and σ_m are statistically significant at the 1% level. Moreover, Table 2 shows that controlling for the endogeneity of the initial conditions in the univariate models is a relevant issue as λ_{11} and λ_{22} are both highly significant with a robust t-ratio of 8.25 and 8.37 respectively.

Results from univariate models in the left panel of Table 2 confirm that laur and lforce are strong predictors of usmigra with F values of 29.16 and 15.68 respectively. Hence, there is strong evidence that these are good candidates to impose exclusion restrictions in the bivariate model. As expected, the unemployment rate on the community's main US destination is detected to have a negative marginal effect on the probability of migration of about 2.5 percentage points (p.p. hereafter). This marginal effect is significantly different from zero at a 1% level. A similar story can be told for the size of the labour force at the community's main US destination. That is, a positive marginal effect of lforce on usmigra of about 1 p.p is found, and results show that such marginal effect is significantly different from zero at 1%.

Migrant network variables have a significant impact on the likelihood of high school graduation. In fact, a Wald test for the joint exclusion of MS_{ji} and $\mathrm{usmigra}_{ji}$ in the univariate model for prepa rejects the null hypothesis at a 1% significance level ($\chi^2(2) = 10.76, \text{p-val} = 0.01$). Moreover, an exclusion test of the proxy for access to community migrant networks, raildis, in dynamic and initial conditions equations for prepa is easily rejected at all standard levels of significance with a $\chi^2(2) = 35.13$, p-val $= 0.00$. Similar observations can be drawn from the univariate model for usmigra.

True dynamic dependance is detected in the univariate model for prepa. In particular, the left panel of Table 2 shows that an increase of 1 unit in the family migrant stock MS_{ji} results in a reduction of about 1.2 percentage points in the likelihood of observing the prepa=1 event. This marginal effect is calculated with high precision. In the case of $\mathrm{usmigra}_{j,i-1}$ a negative marginal effects is also detected. However, in the latter case the marginal effect is obtained with very low precision ($\mid t \mid = -0.34$). Finally, but not least important, a highly significant and positive marginal effect of raildis of around 16% on the likelihood of observing prepa=1 is detected.[12] To put it in other terms, if a community is in a State that was an extra 1000 kilometres farther from the north/south railway in 1900, the model predicts that individuals are 16% more likely to graduate from high school because these individuals have today less access to community migrant networks than individuals in communities that were near the railway by the turn of the 20th century. The negative effects of family and community migration variables detected in the present study are evidence to support the hypothesis that prospective migrants in Mexico who have access to migrant networks behave in a forward looking fashion and drop school early to avoid the wastage of valuable resources.[13]

[12]Notice that a positive effects on raildis is what the analyst should expect because the argument to use raildis says that the farther a community/State was from the railway in 1900 the less developed community migrant networks are in the present day.

[13]Interactions between raildis, and $\mathrm{usmigra}_{j,i-1}$, $\mathrm{prepa}_{j,i-1}$, MS and ES were included in preliminary regressions in both univariate models and in the bivariate model. In all cases, however, such interactions were found to be jointly insignificant. Hence, and to keep a parsimonious specification, the aforementioned interaction terms were excluded from the final specification.

There is also strong evidence from the univariate model of usmigra that true dynamic dependance is present. As expected, the presence of a immediately older family member with US migration experience, $\text{usmigra}_{ji} = 1$, increases the likelihood of usmigra=1 by 3.4 percentage points. In contrast, an extra unit in the migrant stock MS leads to a reduction of about 1.3 percentage points in the likelihood of usmigra=1. Both marginal effects are tightly estimated. While the first observation is consistent with previous research in the migration literature suggesting that the flow of labour across borders generates a self-perpetuating phenomenon due to effects of migrant networks (see, for instance, Delechat 2001, Winters et al. 2001), the latter finding is new and suggests that the self-perpetuating nature of migration have its limits. Intuitively, the negative effect of MS on usmigra may be due to the fact that as more and more members of a family leave the country, the remaining members have to compete less for the resources and assets available to the family in the source country. Further, the larger MS is, the more likely remaining members of the family who stay behind are to be recipients of remittances. As a consequence, they have less incentives to migrate themselves.

A important disadvantage of the univariate models is that $\text{usmigra}_{j,0}$ and MS_{j1} are supposed to be strictly exogenous in the prepa model, and $\text{prepa}_{j,0}$ and ES_{j1} are taken to be strictly exogenous in the usmigra model. As explained in section 4, such assumptions are difficult to met in practise because unobservables at the family level affecting migration and education decisions are likely to be correlated. As a consequence, inconsistent estimators might be obtained. These are important reasons to suggest that a bivariate model is best suited to study the dynamics of education and migration in a coherent framework.

Let us then now move to discuss empirical results from the bivariate model (right panel of Table 2). As before, laur and lforce are highly significant in the migration equations and have their expected signs. Also, estimates for σ_e and σ_m are different from zero at all standard significance levels. Hence, there is strong evidence that intra-family clustering is present in both migration and schooling equations.

A likelihood ratio test for the null of $\rho_u = \rho_v = \rho = \delta_{12} = \delta_{21} = 0$ is provided at the bottom right of Table 2. This is a test for the relevance of the bivariate model over the information already provided by the univariate models. The null hypothesis is easily rejected with a $\chi^2(5) = 46.88$ and a p-val $= 0.00$. Importantly, λ_{12} and λ_{21} are both individually highly significant. This finding implies that the initial conditions equation for usmigra$_{j0}$ is a function of the unobserved random effect f^e and that the initial condition equation for prepa$_{j0}$ is a function of the unobserved random effect f^m. Hence, empirical results show that usmigra$_{j0}$ (MS$_{j1}$) cannot be taken as exogenous in the prepa equation nor prepa$_{j0}$ (ES$_{j1}$) can be taken as exogenous in the usmigra equation.

Like in the two univariate models, family and community migration variables — and therefore, migrant network effects — are found to be strongly significant in both the prepa and usmigra equations (see exclusion Wald test at the bottom right of Table 2). Here, the advantages of estimating the bivariate model over the univariate models are made evident. For instance, we said a few lines above that the marginal effect of MS on prepa=1 in the univariate model is around -1.3 percentage points. In contrast, the marginal probability of MS on the marginal probability of prepa=1 in the bivariate model is -2.4 percentage points. Hence, If one wrongly estimates the univariate model instead of the bivariate model, the marginal effect of MS on prepa would have been underestimated by around 52%. The same specification error would cause the marginal effects of ES on the probability of prepa=1 to be overestimated by 40%. In the case of the usmigra equation results show that wrongly estimating a univariate model leads to the marginal effect of usmigra$_{j,i-1}$ to be underestimated by 12%, the marginal effect of prepa$_{j,i-1}$ by 285%, and the marginal effect of raildis by 53%. In contrast, the marginal effect of MS would have been overestimated by 8%.

Results reported in Table 2 indicate that the correlation between the individual level random terms u^e and u^m, ρ_u, is negative and highly insignificant (robust t-ratio = -5.81). Further, the correlation between the individual level random terms v^e and v^e in the initial condition equations is also found to be negative and significant at the 1%. As a consequence, one can conclude that individual unobservable traits that increase the likelihood of migration are associated as well with reductions in the likelihood of high school graduation. At the family level, in contrast, correlation between unobservable traits f^e and f^m, ρ, is found to be positive. The estimate of ρ is calculated with high precision and the null hypothesis $\rho = 0$ is easily rejected at a 1% level of significance. Hence, data supports the existence of unobservable traits at the family level that increase both the likelihood of usmigra=1 and prepa=1. Clearly, these are valuable empirical findings that only the bivariate random effects model for cluster data can deliver. The apparent puzzle needs careful consideration.

On the one hand negative ρ_u and ρ_v imply that individuals who have a intrinsic ability to study — and do so — migrate less. This finding is consistent with the predictions of the skill sorting model introduced by Borjas (1994) suggesting that Mexican migrants moving to the US are drawn from the bottom tail of the Mexican skills and income distributions, a phenomenon commonly known as "negative migrant selection." On the other hand, a positive ρ suggest that there are unobservable family characteristics that make its members be both more likely to study and more likely to migrate. This suggest some kind of positive migrant selection.

One could speculate that family income and wealth — being two major excluded variables in the regressions due to their likely endogeneity status and a lack of valid instruments for them — might be behind the positive ρ.[14] The argument is as follows. Migration is a costly activity that requires the investment of a minimum quantity of money to pay transportation costs, legal or illegal fees for crossing the border, and settling-up expenses at arrival in the destination country. The poor in Mexico have no access to credit and cannot save the money needed to engage in a emigration venture. It is then people in the middle and top of the income distribution who consider migration to the US as a serious alternative. Importantly, poor people in Mexico will not only find difficult to migrate but also will find strong difficulties to finance their studies in Mexico. As a consequence, given any level of innate individual ability, one expects wealthier families to be better suited to help its members

[14] And speculation is the best the researcher can do at this point because what drives the sign of ρ is correlation between *unobserved* heterogeneity terms at the family level.

to migrate and study than poor families do. From this point of view, finding a positive correlation between unobservables effecting education and migration at the family level ρ is not really a surprise. A similar line of thought has been previously put forward in the literature. In fact, Chiquiar and Hanson (2005) find evidence supporting the hypothesis that migrants come from the middle of Mexico's income and education distribution rather than from the far bottom tail. On the basis of this findings Chiquiar and Hanson challenge the Borjas hypothesis of negative migrant selection, concluding that data suggest rather the presence of intermediate or positive migrant selection. A recent study by Banco de México (2007) gives further evidence to support Chiquiar and Hanson hypothesis.[15]

To put the two pieces of information together one can say that the present study finds evidence that both negative and positive migrant selection in terms of unobservables can co-exist, and that the hypothesis of Borjas and Chiquiar and Hanson are not necessarily mutually incompatible. In the line of Borjas argument, given any level of family income/wealth, unobservable skills at the individual level induce negative migrant selection — i.e., skilled individuals migrate less. Simultaneously, and given any level of individual innate ability, family income/wealth can facilitate both education and migration of the members of a kin and thus induce positive migrant selection. Co-existence of forces selecting positively and negatively migrants leads to the prediction that an independent observer should see most migrants come

[15] In 2005 Banxico ran a survey on migration in 7 border cities: Tijuana, Nogales, Mexicali, Ciudad Jurez, Reynosa, Nuevo Laredo y Matamoros. Among other results, the survey found that most of Mexican migrants were in full employment before departure to the US and had between 6 and 12 years of schooling. This is evidence suggesting that migrants come from the middle of Mexico's education and income distribution.

from the middle rather than from the tails of Mexico's income and education distributions, just as Chiquiar and Hanson and Banco de México report to observe. Hence, the present study somehow bridges apparently contrasting ideas put forward by previous work. The reader, however, should be careful as other explanations of the findings here reported might be correct.

6 Conclusions

The present paper enquires about the potential links between family and community migration and the probability of high school graduation in Mexico. Bivariate random effects dynamic probit models for cluster data are estimated to control for the endogeneity of education and migrant network variables. Correlation of unobservables across migration and education decisions as well as within groups of individuals such as the family are explicitly controlled for. Maximum simulated likelihood techniques are used for estimation.

Findings indicate that an extra migrant in the family decreases the likelihood of high school graduation by 2.4 percentage points. Similarly, individuals who are an extra 1000 kilometres from areas with long-lasting migrant tradition — and have therefore less access to community migrant networks — have increased chances of graduation from high school of about 17.4 percentage points. This evidence supports the hypothesis that prospective migrants in Mexico who have access to migrant networks behave in a forward looking fashion and drop school early to avoid wastage of valuable resources. Estimating a bivariate model instead of two univariate models is of key relevance.

For instance, if the researcher wrongly uses univariate models to analyse the data the aforementioned marginal effects are underestimated by around 52% and 8% respectively.

Regarding migration, empirical evidence suggests that having a migrant in the family increases by 4 percentage points the chances of observing the next younger member of the kin migrating as well. The self-enhancing effects of family migrant networks, however, are found to have a limit because the more members of the kin migrate the less likely the remaining members are to migrate themselves. Community migrant networks are found to increase the odds of migration. In fact, evidence shows that being an extra 1000 kilometres nearer from areas with long-lasting migrant tradition increases the likelihood of migration by 9 percentage points. Wrongly using univariate models for migration in place of the bivariate model leads to an underestimation of the aforementioned marginal effects by 12% and 53% respectively.

Negative and positive migrant selection in terms of unobservables are both supported by the data. Negative migrant selection is found at the individual level while positive migrant selection is detected at the family level. As suggested by Borjas (1994), the negative migrant selection at individual level is likely to indicate that, given family income and wealth, skilled individuals tend to migrate less and study more. However, given any level of individual innate ability and because migration and education are costly activities that poor households have limited access, the paper speculates that wealthier families are better suited to finance both migration and education of its members. As a consequence, differences in unobserved — uncontrolled — family wealth can be capable of creating positive migrant selection. Co-existence of forces selecting positively and negatively migrants leads to the prediction that an independent observer should see most migrants come from the middle rather than from the tails of Mexico's income and education distributions. This latter explanation is line with the findings of Chiquiar and Hanson (2005) showing that Mexican migrants are mainly drawn from the middle of Mexico's income and education distributions and that it seems to be intermediate or positive migrant selection rather than negative selection in the Mexico-US flows of labour. Hence, the present paper gives hints of how negative and positive migrant selection may co-exist simultaneously and help to bridge apparently contradicting ideas and findings on migration selectivity put forward by previous work.

Table 1. Descriptive Statistics

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
Individual characteristics						
sex	=1 if female	11990	0.51	0.50	0	1
age	age in years	11990	33.13	11.15	18	60
usmigra	=1 if ever migrated to the US	11990	0.18	0.39	0	1
prepa	=1 if completed high school	11990	0.21	0.40	0	1
Head of household						
hsex	=1 if female	11990	0.12	0.32	0	1
hage	age in years	11990	46.25	10.02	18	60
hnchild	No. of children ever born	11990	4.83	2.87	0	14
Migration/Education stock at family level						
MS	Migrant stock	11990	0.40	0.84	0	9
ES	Education stock	11990	0.28	0.75	0	7
State/Community						
urban	=1 if urban community	11990	0.49	0.50	0	1
raildis	Distance to railway in 1900 (000s km)	11990	0.12	0.20	0	1.46
laur	Unemployment rate (%) in main US destination	11990	6.18	1.27	4.10	13.64
lforce	Labour force (millions) in main US destination	11990	41.38	21.33	0.74	85.25
Birthplace (Base Mex. City and environs)						
North	North	11990	0.30	0.46	0	1
Centre	Centre	11990	0.22	0.42	0	1
CentreP	Centre Pacific	11990	0.16	0.37	0	1
South	South	11990	0.17	0.37	0	1
Survey year (Base 1997)						
yr1998	1998	11990	0.25	0.43	0	1
yr1999	1999	11990	0.08	0.27	0	1
yr2000	2000	11990	0.10	0.31	0	1
yr2001	2001	11990	0.14	0.35	0	1
yr2002	2002	11990	0.08	0.27	0	1
yr2003	2003	11990	0.12	0.33	0	1
yr2004	2004	11990	0.15	0.35	0	1

Table 2. Random effects dynamic Probit results — Marginal effects

	Univariate models				Bivariate Model			
	prepa		usmigra		prepa		usmigra	
Variable	ME	$\mid t \mid$	ME	$\mid t \mid$	ME$^{(a)}$	$\mid t \mid$	ME$^{(a)}$	$\mid t \mid$
Individual characteristics								
sex	0.008	1.40	-0.114	-8.02	0.010	1.51	-0.120	-8.52
age	-0.006	-12.20	-0.002	-5.58	-0.006	-12.60	-0.002	-5.64
Head of household								
hsex	-0.003	-0.30	0.009	0.72	-0.005	-0.41	0.008	0.67
hage	0.006	8.97	0.001	3.07	0.006	9.52	0.002	3.59
hnchild	-0.023	-8.48	0.006	4.18	-0.024	-9.13	0.006	3.55
lagged dependent variables								
usmigra$_{j,i-1}$	-0.003	-0.34	0.034	3.25	-0.017	-1.81	0.038	3.39
prepa$_{j,i-1}$	0.074	4.20	0.010	0.94	0.082	4.17	-0.005	-0.49
Migration/Education stock								
MS	-0.012	-2.45	-0.015	-4.29	-0.024	-3.36	-0.014	-3.68
ES	-0.015	-3.11	-0.013	-2.51	-0.011	-1.73	-0.026	-3.33
State/Community								
urban	0.070	6.65	-0.084	-2.38	0.075	6.94	-0.041	-4.56
raildis	0.159	4.92	-0.042	-4.80	0.174	4.90	-0.090	-2.39
laur			-0.025	-5.04			-0.026	-4.49
lforce			0.001	3.96			0.001	3.87
Birthplace and year dummies								
Birthplace	yes		yes		yes		yes	
Year	yes		yes		yes		yes	
Auxiliary parameters	Coeff.	$\mid t \mid$	Coeff.	$\mid t \mid$			Coeff.	$\mid t \mid$
λ_{11}	0.77	8.25					0.74	7.61
λ_{12}							-0.18	-2.54
λ_{21}							-0.26	-2.29
λ_{22}			0.68	8.37			0.86	7.30
σ_e	1.22	10.82					1.14	10.23
σ_m			1.03	10.86			0.98	11.01
ρ_u							-0.34	-5.81
ρ_v							-0.21	-2.28
ρ							0.34	3.02
Exclusion Wald tests								
Edu vars.$^{(b)}$		50.17 (0.00)		7.88 (0.02)		46.87 (0.00)		16.08 (0.00)
Family Migr. vars.$^{(c)}$		10.76 (0.01)		29.77 (0.00)		21.96 (0.00)		28.91 (0.00)
Cummunity Migr. vars.$^{(d)}$		35.13 (0.00)		12.00 (0.01)		34.99 (0.00)		12.35 (0.01)
Model relevance								
$\rho_u = \rho_v = \rho = \lambda_{12} = \lambda_{21} = 0$					$\chi^2(5) = 46.88$ (pval = 0.00)			
Model information								
No. Halton draws		400		400				400
No. families		3479		3479				3479
No. observations		11990		11990				11990
Log-likelihod		-4627.6		-4666.3				-9262.7

Note. Marginal effects are evaluated at the means of the explanatory variables. For dummy variables, they show the change in the relevant probability when the variable changes from 0 to 1. Robust t-ratio $\mid t \mid$ for marginal effects are reported. $^{(a)}$ Marginal effects on marginal probabilities. $^{(b)}$ Joint test for exclusion of the education dummies in dynamic equations. This is a test for the exclusion of: ES and prepa$_{j,i-1}$ (p-values in brackets). $^{(c)}$ Joint exclusion test of family migration variables in dynamic equation. This is a test for the exclusion of: MS, and usmigra$_{j,i-1}$ (p-values in brackets). $^{(d)}$ Joint exclusion test of raildis in dynamic and initial conditions equations. Results from initial conditions are available from the author upon request.

References

Alessie, R., Hochguertel, S., Van Soest, A., 2004. Ownership of stocks and mutual funds: A panel data analysis. The Review of Economics And Statistics 86, 783–796.

Arulampalam, W., Bhalotra, S., 2006. Sibling death clustering in india: Genuine scarring vs unobserved heterogeneity. The Journal of the Royal Statistical Society, Series A (forthcoming).

Banco de México, 2005. Informe anual.

Banco de México, 2007. Las remesas familiares en mexico: inversin de los recursos de migrantes.

Blundell, R., Dearden, L., Goodman, A., Reed, H., 2000. The returns to higher education in Britain: Evidence from a british cohort. Economic Journal 110 (461), 82–100.

Borjas, G., 1994. The economics of immigration. Journal of Economic Literature 32 (4), 1667–1717.

Chiquiar, D., Hanson, G. H., 2005. International migration, self-selection, and the distribution of wages: Evidence from mexico and the united states. Journal of Political Economy 113 (2), p239 – 281.

CONAPO, 2000. Migracion Mexico-EU presente y futuro. CONAPO. URL http://www.conapo.gob.mx/

Delechat, C., 2001. International migration dynamics: The role of experience and social networks. Labour 15 (3), 457–486.

Durand, J., Kandel, W., Parrado, E., Massey, D., 1996. International migration and development in mexican communities. Demography 33 (2), 249–264.

Goldstein, H., 2003. Multilevel statistical models, 3rd Edition. Kendall's Library of Statistics. Arnold, London.

Gourieroux, C., Monfort, A., 1993. Simulation-based inference: A survey with special reference to panel data models. Journal of Econometrics 59, 533.

Hanson, G., Woodruff, C., 2003. Emigration and educational attainment in Mexico, University of California at San Diego, manuscript.

Heckman, J., 1981a. Heterogeneity and state dependence. In: Rosen, S. (Ed.), Studies in Labor Markets. University of Chicago Press, pp. 91–139.

Heckman, J., 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, pp. 179–195.

Heckman, J. J., 1978. Dummy endogenous variables in a simultaneous equation system. Econometrica 46 (4), 931–959.

Keane, M. P., 1992. A note on identification in the multinomial probit model. Journal of Business & Economic Statistics 10 (2), 193–200.

McCulloch, C., Searle, S., 2001. Generalized, linear,and mixed models. Wiley Series in Probability and Statistics. John Wiley & Sons, INC, New York.

McKenzie, D., Rapoport, H., 2006. Can migration reduce educational attainment? evidence from Mexico, World Bank policy research working paper No. 3952.

Miranda, A., Bratti, M., 2006. Non-pecuniary returns to higher education: The effects on smoking in the UK, IZA Discussion Paper No. 2090.

Taibo II, P., 2006. Pancho Villa: Una Biografia Narrativa. Editorial Planeta.

Train, K., 2003. Discrete choice methods with simulation. Cambridge university press.

UNPD, 2006. International migration wall chart 2006, United Nations Population Division.

Vidal, J.-P., 1998. The effect of emigration on human capital formation. Journal of Population Economics 11 (4), 589–600.

Winters, P., De Janvry, A., Sadoulet, E., 2001. Family and community networks in Mexico-U.S. migration. Journal of Human Resources 36 (1), 159–184.

Woodruff, C., Zenteno, R., 2007. Migration networks and microenterprises in mexico. Journal of Development Economics 82 (2), 509–528.

World Bank, 2006. Economic implications of remittances and migration, Global Economic Prospects.